R E P O R T  R E S U M E S

ED 003 585 56

PRINCIPLES OF PROGRAMING MATERIAL FOR TEACHING MACHINES AND THEIR RELATION TO TRANSFER OF TRAINING.

BY- LEVIN, GERALD R.
BROWN UNIV., PROVIDENCE, R.I.

REPORT NUMBER NDEA-VIIA-538 PUB DATE 61
GRANT OEG-7-37-088
EDRS PRICE MF-$0.09 HC$0.96 24P.

DESCRIPTORS- *TEACHING MACHINES, *PROGRAMING, *GRADE 2, *GEOMETRY, AUTOINSTRUCTIONAL METHODS, *AUTOINSTRUCTIONAL PROGRAMS, ACHIEVEMENT TESTS, PROGRAMED UNITS, PROVIDENCE, RHODE ISLAND

THE PURPOSE WAS THE DISCOVERY AND EVALUATION OF GENERAL PROGRAMING PRINCIPLES FOR TEACHING MACHINES. THE DEVELOPMENT OF A SELF-INSTRUCTIONAL PROGRAM IN INFORMAL GEOMETRY FOR SECOND-GRADERS WAS THE FIRST STEP OF THE PROCEDURE. THE 170-ITEM PROGRAM CONSISTED OF FIVE UNITS DESIGNED (1) TO FAMILIARIZE THE CHILD WITH MACHINE OPERATION AND (2) TO TEACH SUCH CONCEPTS AS THE STRAIGHT LINE, THE ANGLE, AND THE TRIANGLE. ACHIEVEMENT TESTS WERE DEVELOPED TO ASSESS THE EFFECTS OF THE SELF-INSTRUCTIONAL PROGRAM, AND TO FACILITATE THE SELECTION AND DESCRIPTION OF RESEARCH SUBJECTS. FIFTEEN PUPILS WERE TAKEN THROUGH THE PROGRAM AND COMPARED WITH MATCHED CONTROLS WHO DID NOT GO THROUGH THE PROGRAM. A STUDY WAS CONDUCTED TO CHECK THE ASSUMPTION THAT THE KNOWLEDGE OF GEOMETRY COVERED IN THE CHILDREN'S PROGRAM VARIED LITTLE DURING THE PRIMARY GRADES. A STUDY WAS ALSO CONDUCTED TO DETERMINE THE IMPORTANCE OF PRESENTING PROGRAM ITEMS IN THE ORIGINAL AND LOGICAL ORDER IN WHICH THEY HAD BEEN DEVELOPED. MATCHED GROUPS OF 18 SUBJECTS EACH WERE TAKEN THROUGH TWO VERSIONS OF THE PROGRAM. THE CONTROL GROUP WENT THROUGH THE STUDENT PROGRAM, WHILE THE EXPERIMENTAL GROUP WENT THROUGH A VERSION OF THE PROGRAM IN WHICH THE ITEM OF ONE UNIT HAD BEEN RANDOMLY SCRAMBLED WITHIN BLOCKS OF 20 ITEMS. THE LEARNING ACHIEVEMENT OF THE EXPERIMENTAL GROUP WAS NOT SIGNIFICANTLY DIFFERENT FROM THAT OF THE CONTROL GROUP. (HB)

ED003585

S-38

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

# Principles of Programming Material for Teaching Machines and their Relation to Transfer of Training

Gerald R. Levin

Title VII Project Number 533

National Defense Act of 1958

Grant Number 7-37-088

Brown University

The Research Reported Herein Was

Supported by a Grant from the

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education

ED003585

The research project was oriented toward the discovery and evaluation of general principles for programming material for teaching machines. Specifically, the research was aimed at the investigation of such questions as, "How important is it to order the items in a teaching machine program with care?"

The present report includes two sections 1) a brief summary of the general accomplishments of the research project, and 2) a report of the final experiment which was carried out in order to examine the importance of the item sequence variable.

## Summary of Accomplishments

I. Program development. The first stage of the research program was the development of a self-instructional program in informal geometry for second graders. Geometry was selected as the subject to be programmed because of 1) its logical structure; 2) its suitability for presentation through a printed medium; 3) its assumed similarity to other fields of mathematics in terms of the kinds of learning involved; and 4) its unfamiliarity to children in the primary grades. A program was developed consisting of five units designed to familiarize the child with machine operation, and to teach such concepts as straight line, angle, and triangle, along with appropriate notation. The program consisted of 170 items and was developed for presentation by means of Foringer #2002 Teaching Machines. A full description of the program, including copies of all items, can be found in Levin, 1961a.

II. Test Construction. One of the central concerns in the research program was whether popular principles of programming, e.g., minimize error rate during acquisition, would turn out to be important if retention and transfer criteria were employed to evaluate learning. Accordingly, achievement tests were developed to assess the effects of the self-instructional program, and to facilitate the selection and description of research subjects. Specimen sets of the tests developed can be found in Levin, 1961b.

III. Program validation. A study was conducted in which 15 second graders were taken through the geometry program and compared with matched controls who did not go through the program. A comparison of gain scores on achievement tests indicated that the program reliably produced an increased knowledge of the material covered. However, the study also indicated that the program was not markedly effective in terms of how well program goals were achieved. This study is summarized in Levin, 1961a and 1961b.

IV. Geometry knowledge and grade level. A study was conducted to check on the assumption that children's knowledge of the geometry covered on the program varied little during the primary grades. Children from two second grades, one fourth, and one sixth grade were compared with respect to their performance on geometry tests. The results indicated a slight increase in test scores as a function of grade level. However, these increases were

small in comparison to those produced by taking children through the program, or even by retesting children without any intervening instruction. This study is summarized in Levin, 1961b.

V. Conceptions of triangularity. Achievement test data indicated that children's misconceptions about triangles displayed striking regularity. An ancillary study was conducted to explore this problem more systematically. The data supported the assumption that uninstructed children with partial knowledge of what a triangle was conceived of a triangle as an isosceles triangle with vertical angle about 50 degrees and base parallel to the horizontal. This study will be submitted to Child Development.

VI. Effects of item scrambling. A study was conducted to determine the importance of presenting program items in the original, logical order in which they had been developed. Matched groups of 18 Ss each were taken through two versions of the program. The control group went through the standard program, while the experimental group went through a version of the program in which the items of one unit had been randomly scrambled within blocks of 20 items. Evaluations of acquisition, retention, and transfer criteria of learning all indicated that the experimental treatment had no effects on learning. A full report of this study can be found below.

## Staff

Bruce L. Baker. Undergraduate research assistant. September, 1960 through June, 1961.

Anthony Hilton. Undergraduate research assistant. June, 1961.

Neil Maurer. Undergraduate research assistant. June, 1961.

Susan Simon. Undergraduate research assistant. Miss Simon was supported by an NSF undergraduate fellowship while working on the project. September, 1960 through May, 1961.

Gerald R. Levin. Principal Investigator. June, 1960 through June, 1961.

## References

Levin, G.R. Informal geometry for the second grade. Part 1. An instructional program. Providence, R.I.: Brown University, 1961a.

Levin, G.R. Informal geometry for the second grade. Part 2. Measuring achievement. Providence, R.I.: Brown University, 1961b.

# Item Scrambling in a Self-instructional Program[1]

Gerald R. Levin

&

Bruce L. Baker

Brown University

Decisions about the sequence in which material is presented are inherent in the process of constructing self-instructional programs. In a program designed to teach several concepts, one concept must be selected for initial treatment. If ten items are used to teach a concept, these items must be ordered. While every programmer must make assumptions about ordering topics and items, there is little experimental evidence available which bears on the validity of such assumptions.

Gavurin and Donahue (1961) apparently have conducted the only experiment directly relevant to the question of item sequence. They used the first 29 items of a program on psychology (Holland and Skinner, 1961) and presented them in a standard sequence to one group of adults, and to another group in a random sequence. Their method of scrambling items was a combination of what might be termed block scrambling and item scrambling. The twenty-nine items were divided into blocks of ten, ten, and nine items. The order of blocks was scrambled for the experimental group. Within each block, the nine or ten items also were scrambled so that the final random order contained more local homogeneity than would have been expected if scrambling had taken place within the total pool of 29 items. The Ss were required to reach a criterion of an errorless trial on all nine or ten items within a block before moving on to the next block. It was found that the group which went through the items in a random sequence required significantly more trials to criterion and made more errors during acquisition. However, when retention was tested a month later, the superiority of the group which went through the standard sequence of items was not statistically significant. Unfortunately, the design of their study did not permit a comparison between the two treatments with the number of trials held constant.

The present experiment, like the study by Gavurin and Donahue, was conducted to determine the importance of presenting items in a standard, logical sequence which had been arrived at on the basis of prior planning and experimentation. The original impetus to the study was a concern that if a program were eventually to be evaluated in terms of criteria of retention and transfer, then error rate during acquisition might not be a satisfactory criterion to use while developing a program. Accordingly, the random presentation of program items appeared to be a convenient means of varying error rate during acquisition while holding, other program characteristics constant. The focus of the study then, was to see what happened to varied criteria of learning when item sequence was varied in a portion of a self-instructional program.

## Method

A matched-group experiment was planned which would make it possible to examine the effects of scrambling item sequence in one unit of a program on (1) error rate during acquisition, (2) performance on a subsequent unit of the program, and (3) performance on an achievement test which measured retention and transfer.

### Subjects

The 36 Ss were all pupils in the last month of second grade in a public school. These 13 boys and 23 girls were selected from among 81 children who were present in three classes on the day a screening test, Test S, was administered. The final sample was selected from among those who met the initial, screening criteria by subsequently dropping five additional children. Three children were dropped because of poor performance on the first two units of the program. Subsequently, one child in each experimental group was dropped

because of procedural errors during the experimental treatment. Thus, the final sample consisted of two groups of 18 Ss each.

Both the experimental and control groups had mean ages of 7.6 years with standard deviations of 0.6 years. The mean IQ (Otis tests) of the control group was 107.0 with a standard deviation of 11.3, while for the experimental group the corresponding figures were 107.4 and 14.4

Instruments

The research instruments included two forms of a specially developed self-instructional program, teaching machines for presenting the programs, and two tests.

Programs. The self-instructional program, Informal geometry for the second grade, had been developed through a process of trial and revision as suggested by Skinner (1958). A validation study had established that the program produced a statistically significant amount of learning, but was not highly effective in promoting masters of the material.

The program consisted of 180 items, or steps, grouped in five units: Alpha, Dots, Lines, Angles, and Triangles. Alpha was developed to use while teaching children how to manipulate the teaching machines and did not contain any geometry items. The other four units comprised a linear sequence in which concepts and notation taught in one unit were used and built upon in subsequent units. The program had a mean of 6.7 words per item. Only 99 different words were used in the program and the Type Token Ratio for the program as a whole was .82. The 15 second graders who went through the program in the validation study had a median error rate of 22.2% for the total program and took a median time of 59.5 minutes to go through the items on their first exposure to them.

The only portion of the experimental program which differed from the standard form was the unit on angles. The angles unit was selected for experimental treatment because of its difficulty and because of its position in the program. Previous results indicated that the angles unit was the most difficult. Moreover, the fact that the unit on angles came midway in the program made it possible to compare the treatment groups with respect to acquisition before and after the experimental treatment, as well as during it.

In the standard program the unit on angles consisted of 60 items which were presented on three strips, each containing 20 items. In the experimental form of the program the order of these three blocks of 20 items was held constant, while the order of the 20 items within each block was randomly scrambled. The rank order correlations between standard and random item sequences in the three blocks were -.17, -.28, *and -.07 respectively.*

Program Presentation. The program was duplicated on 8 1/2 x 42 inch strips of paper by the ozalid process and these strips were presented by means of Foringer #2002 Teaching Machines. The items were spaced so that it was impossible to see portions of two items at once. The children wrote their composed answers on a standard answer strip.

Tests. Two achievement tests had been developed for different purposes. Test S was designed for quick screening purposes. It was designed to screen out Ss whose verbal comprehension was too poor for them to understand the program, and Ss who were well acquainted with the material to be covered in the program. Test S consisted of two mimeographed pages

containing eight verbal comprehension items and eight geometry items. Test 1 was developed to provide a means of assessing what a child learned by going through the program. The 52 items were grouped homogeneously on seven mimeographed pages.

Procedure

The experiment began when the screening test was administered to the children in their classrooms. It ended 17 days later. During the experiment every S participated for about 15 minutes every day with only minor exceptions.

After the initial screening, all subsequent research sessions were carried out in a psychological laboratory in the basement of a nearby school, School$^{0}$. Most of the Ss had gone to first grade in School$^{0}$ and it was common for pupils in their building to go to School$^{0}$ on errands, etc.

The Ss typically went from their classrooms to the laboratory in groups of four. Members of the experimental and control groups came to the laboratory simultaneously and in approximately equal numbers at all stages of the experiment. Two or three experimenters were in the laboratory at all times, and they timed each child on each program segment with stop watches.

Pretesting and Pretraining. In their first session in the laboratory, Ss were instructed, given the first five pages of Test 1, and taught to use the teaching machines while going through Unit Alpha of the self-instructional program. The next day pretraining began. During pretraining each S went through the units on dots and on lines twice. After

completing the second trial on lines, each S was retested with three pages of Test 1 which covered the units on Dots, Lines, and Angles.

Experimental treatment. Ss were matched and assigned randomly to treatment groups on the basis of their scores on the angles items given during the first administration of Test 1. The preferred procedure of matching them on the basis of scores on both administrations of Test 1, IQ, and performance during pretraining had to be foregone due to time pressures.

There were no cues except for those provided in the program to indicate to Ss that the procedure was different during the experimental treatment than previously. Throughout the experiment, Ss were told that different children were working on different material.

During the experimental treatment each S went through each segment of the angles program three times before moving on to the next segment.

Posttraining and Posttesting. After they finished the angles unit, all Ss were taken through the final unit of the program, Triangles, presented in standard form. All of the Ss finished the program on a Friday. On the following Monday they all were given Test 1 and Test S.

Data to be Examined

In order to determine the importance of the item sequence variable, several comparisons were necessary. Comparisons of the experimental and control groups with respect to pretest and pretraining performance would determine whether they were performing similarly before the introduction of the experimental treatment. An analysis of speed and error data during the experimental treatment would then be required to determine whether

the treatment was related to acquisition performance. Performance on the Triangles unit of the program, scores on posttests, and comparisons of pretests and posttests would indicate whether retention and transfer criteria of learning showed effects of the experimental treatment. Finally, correlations between IQ, pretest scores, program scores, and posttest scores for each group would reveal how response measures interrelated, and whether such correlations would clarify group comparisons.

## Results

The data indicated that program performance throughout the experiment was similar for the experimental and control groups. Table 1 summarizes time and error score for pretreatment, treatment,

.....

Insert Table 1 about here

.....

and posttreatment performance on the program. Mann-Whitney U Tests indicated that none of the differences in Table 1 reached statistical significance. Figures 1 and 2 facilitate a more

.....

Insert Figure 1 about here

.....

detailed examination of program performance. These figures bear out the grosser comparisons in Table 1.

.....

Insert Figure 2 about here

.....

Table 2 makes it possible to see the effects of the experimental treat-

.....

Insert Table 2 about here

.....

ment on both specific and general measures of knowledge of geometry. Mann-Whitney U Tests indicated that none of the comparisons approached statistical significance. Additional comparisons of such facets of test performance as transfer items, non-transfer items, and individual pages of the posttest also failed to indicate any differences in performance between the two groups.

Thus, group comparisons of acquisition, retention, and transfer data all indicated that the experimental treatment did not have statistically significant effects.

Correlations were computed for several response measures to clarify the meaning of the group comparisons. In general it was found that measures of intelligence and quality of performance in geometry tended to be correlated consistently, while measures of speed of performance and gain from pretests to posttests were less consistent. Table 3 summarizes the correlations for intelligence and three, gross measures of geometry knowledge.

.....

Insert Table 3 about here

.....

## Discussion

While the meaning of the results is, of course, subject to varied interpretation, the data certainly provide no support for the contention that item sequence is an importance variable in self-instructional programs. There was no evidence that item scrambling impaired learning. Moreover, the most reliable of the many, statistically insignificant comparisons was in favor of the experimental group.

The correlations between measures of individual differences at the beginning of the experiment with indices of learning suggests that individual differences might have obscured treatment effects. Analysis of the present data also indicated a slight tendency for bright children in the control group to do better than bright children in the experimental group, while the opposite trend appeared for the duller children. Thus, it seems clear that the possibility of interaction between initial ability and treatment effects deserves explicit investigation. However, it should be noted that the Ss in the present sample probably were no more variable with respect to intelligence than the children in an average, American classroom. Thus, individual differences probably would be as important in many classroom settings as in the present experiment.

Probably the most important limitation of the present study was the failure of the program to teach the material presented thoroughly to most of the Ss, as indicated in posttest performance. It seems likely that a more effective program would have been made less effective when scrambled.

While the present findings failed to support the assumption that

item sequence is important, it seems neither appropriate, nor even tempting to abandon the hypothesis that the order of presentation matters under some conditions. Instead, it seems useful to try and conceptualize the problem more adequately and operationally.

It is intriguing to think of item scrambling as a means of dissecting a program so as to learn more about its sequential structure. When we hold the sequence of units, or blocks of items, constant and vary item order within blocks, as in the present experiment, we are in a position to learn about the "fine structure" of the program. When we re-order blocks of items and hold order constant within these blocks, we can explore the "gross anatomy" of the program. A program consisting of totally unrelated items presumably would be found to have "no sequential structure," as indicated by no differences in indices of learning, regardless of what we did to the order of items. In contrast, a program with "completely linear structure" would be less effective if we tampered with its sequential structure in any manner, or to any degree.

It seems likely that such dissection would reveal that most existing programs are neither completely linear, nor completely lacking in sequential structure. Moreover, such dissection might lead to a useful means of describing and comparing the sequential structures of programs. Finally, we may only know how much emphasis to put upon sequence in programming when we can empirically characterize the sequential structure of self-instructional programs and have sound empirical generalizations about the relationships between sequential structure, content, and dependent variables.

## Summary

A study was made of the importance of item sequence in a self-instructional program in geometry. Thirty-six second graders were divided into two matched groups and taken through two different versions of the same program. The control group went through the standard program while the experimental group went through most of the standard program and a unit of the standard program in which the sequence of items had been randomly scrambled. The groups were compared with respect to time and error scores while going through the program, and on achievement tests related to program content. It was found that there were no statistically significant differences between the two groups on indices of acquisition, retention, or transfer. While the results made it necessary to question the importance of item sequence in self-instructional programs, it was felt that 1) a possible interaction between individual difference measures (e.g., IQ) and self-instruction performance exists, and 2) there may be a relationship between over-all program effectiveness and ability to benefit from a non-random sequence of items.

## References

Gavurin, E.I., & Donahue, Virginia M. Logical sequence and random sequence. Automated Teaching Bull., 1961, 1, 3-9.

Holland, J.G., & Skinner, B.F. The analysis of behavior. New York: McGraw-Hill, 1961.

Skinner, B.F. Teaching machines. Science, 1958, 128, 969-977.

Footnotes

[1]The research reported herein was supported in part by Grant No. 7-07-088.00 from the United States Office of Education, Department of Health, Education, and Welfare.

The authors are indebted to Edward Fitzgerald, Superintendent of the Bristol, R.I. School Department for his cooperation in facilitating this research. Anthony Hilton's and Neil Maurer's assistance in the collection and analysis of data is gratefully acknowledged.

Table 1

Program Performance

| Stage of Training | Median Errors | | | Median Working Time (Minutes) | | |
|---|---|---|---|---|---|---|
| | C[a] | E[b] | T[c] | C[a] | E[b] | T[c] |
| Pretreatment | 13.5 | 18.5 | 313.0 | 32.2 | 32.8 | 327.5 |
| Treatment | 47.5 | 50.5 | 321.0 | 46.9 | 49.5 | 332.0 |
| Posttreatment | 16.5 | 19.0 | 308.0 | 17.9 | 14.8 | 282.0 |

[a]Control group.

[b]Experimental group.

[c]Smaller rank total. A rank total of 333.0 would indicate identical sums of ranks for the two groups. A rank total of 271.0 would provide a $p$ value of .05.

Table 2

Test Performance

| Test Measure | Group | | Smaller Rank Total[a] |
|---|---|---|---|
| | Control Median | Experimental Median | |
| Angles Pretest 1 | 2.5 | 2.0 | 332.5 |
| Angles Pretest 2 | 3.0 | 3.0 | 308.5 |
| Angles Posttest | 8.0 | 7.0 | 310.5 |
| Angles Gain | 4.2 | 5.0 | 326.5 |
| Total Pretest 1 | 31.0 | 29.2 | 306.0 |
| Total Posttest | 50.2 | 47.0 | 316.0 |
| Total Gain | 18.0 | 18.2 | 333.0 |

[a]A rank total of 333.0 would indicate identical sums of ranks for the two groups. A rank total of 271.0 would have a P value of .05.

Table 3

Selected Correlations

(N = 18 for each group)

| Variables[a] | Group: Control | Group: Experimental |
|---|---|---|
| IQ & Pretest | .24 | .50* |
| IQ & Program | .68** | .58** |
| IQ & Posttest | .45* | .56** |
| Pretest & Program | .52* | .48* |
| Pretest & Posttest | .52* | .65** |
| Program & Posttest | .80*** | .51* |

[a]Pretest, posttest, and program scores refer to number of items answered correctly.

*Significant at the .05 level, with a one-tailed test.

**Significant at the .01 level, with a one-tailed test.

***Significant at the .0005 level, with a one-tailed test.

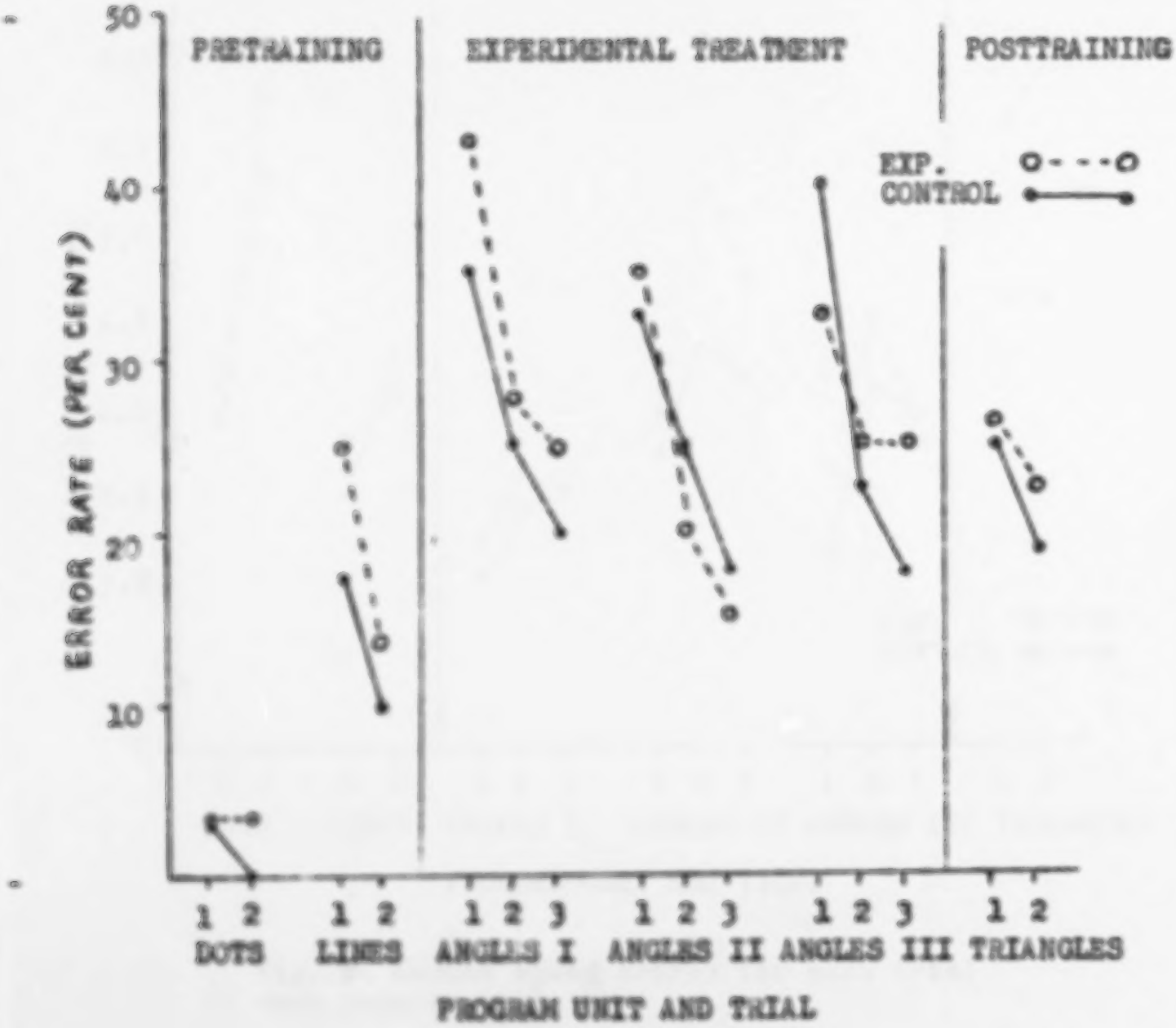


Fig. 1. Median error rates for each trial on each program unit.

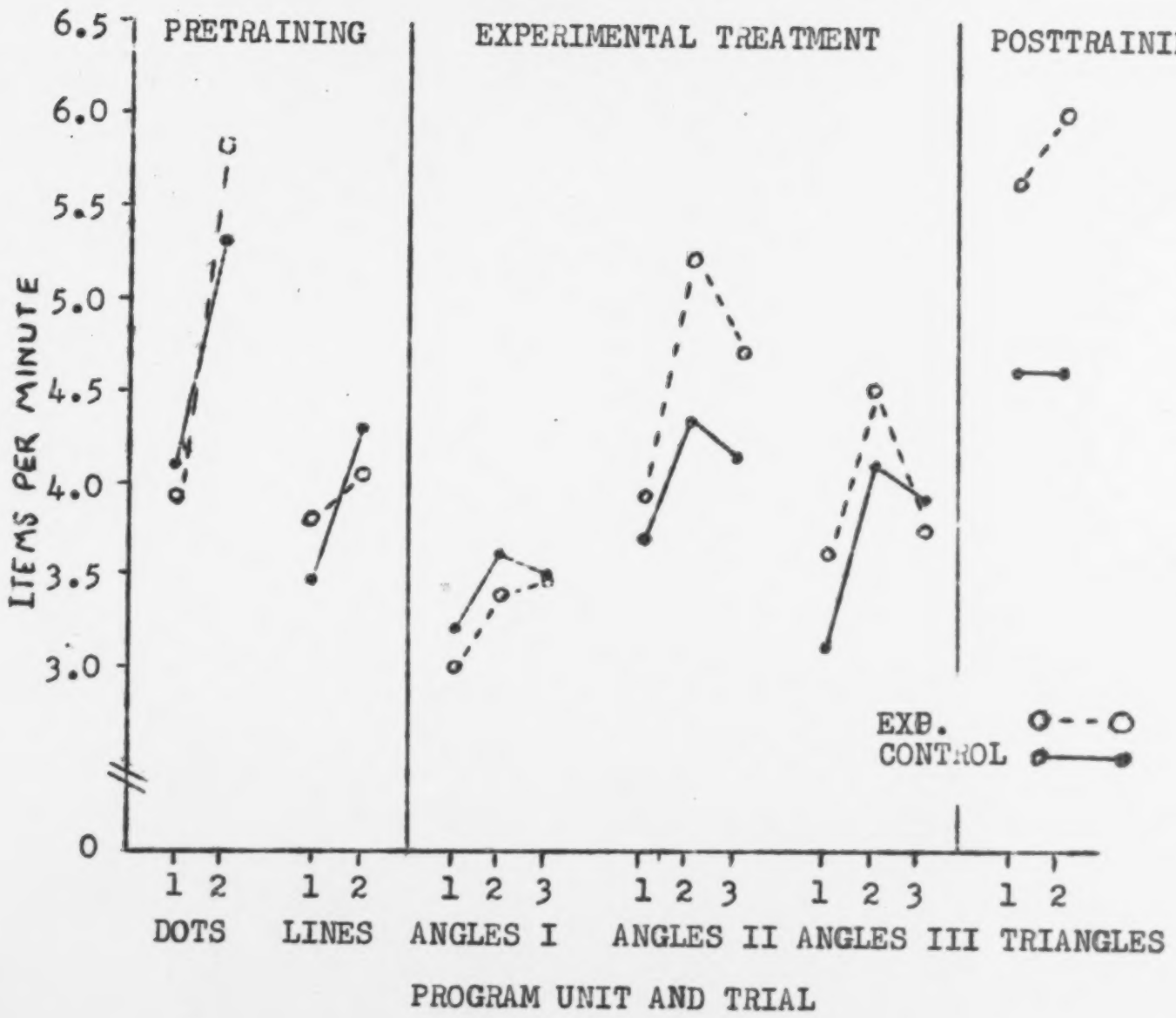


Fig. 2. Median speed scores for each trial on each program unit.

ITEM 1